Disclaimer: This is not the final version of the article. Changes may occur when the manuscript is published in its final format.

Antipsychotic Prescribing and Delirium Outcomes at a Mexican Tertiary Referral Center: A Retrospective Cohort

Alejandro Salazar-Rodríguez 1, Alejandro Díaz-Bernal 1,*, Deldhy Nicolás Moya-Sánchez 1, Marta Georgina Ochoa-Madrigal 1, Oscar Meneses-Luna 1, Elia Francisca Espinoza-de-Ávila 1 and Francisco Javier Cruz-Aviña 1

1 Department of Psychiatry, Consultation-Liaison Psychiatry, and Child and Adolescent Psychiatry, Centro Médico Nacional "20 de Noviembre," ISSSTE / Facultad de Medicina, UNAM. Mexico City, Mexico.

Author e-mail addresses: Alejandro Salazar-Rodríguez [asalazar961@gmail.com]; Alejandro Díaz-Bernal [alejopoa@gmail.com]; Deldhy Nicolás Moya-Sánchez [Deldhym@gmail.com]; Marta Georgina Ochoa-Madrigal [marginocho@gmail.com]; Oscar Meneses-Luna [meneses.luna.72@gmail.com]; Elia Francisca Espinoza-de-Ávila [dra_eliaespinoza@yahoo.com]; Francisco Javier Cruz-Aviña [paco.avina22@gmail.com].

** Correspondence: alejopoa@gmail.com* | *RPI: 643.2024*

Abstract

Background: Haloperidol remains the default antipsychotic for hospital delirium in much of the world, yet prescribing practice and measured outcomes have rarely been described together in Latin American hospitals.

Methods: Retrospective cohort of consecutive adults with delirium seen by consultation-liaison psychiatry at a Mexico City tertiary referral centre (March 2021 to August 2022). Severity was the summed CAM-S short-form score (0–7) at each psychiatric visit; each patient contributed one index antipsychotic regimen. We modelled CAM-S with linear mixed models and estimated cumulative incidence of first resolution (CAM-S = 0) with the Aalen–Johansen estimator, treating in-hospital death as a competing event. No comparative-effectiveness hypothesis was prespecified.

Results: Of 195 screened records, 136 patients contributing 724 assessments were analysed; haloperidol accounted for 115 regimens (84.6%). CAM-S fell by 0.38 points per day (95% CI 0.30–0.45; $p < 0.001$), from 3.79 to 2.21. Adjusted for baseline severity, no agent differed from haloperidol (olanzapine −0.50, 95% CI −1.04 to 0.05; risperidone −0.23, −0.69 to 0.24). Cumulative incidence of first resolution reached 38.3% (95% CI 27.1–49.4) by day 7 and 46.1% (31.8–59.2) by day 14. Baseline CAM-S was the only variable associated with resolution (hazard ratio 0.49 per point, 0.36–0.67). Physician-documented improvement at discharge (77.9%) far exceeded instrument-based severity control (25.0%).

Conclusions: Prescribing was concentrated on haloperidol while measured severity improved independently of which agent was chosen. The gap between charted improvement and measured outcomes reflects two different constructs, not a treatment effect. These data describe local practice and cannot support causal comparison between agents.

Keywords: delirium; haloperidol; olanzapine; confusion assessment method; consultation-liaison psychiatry; antipsychotic prescribing.

1. Introduction

Delirium affects 14–56% of hospitalised patients, with a pooled prevalence of 23.6% (95% CI 19–29%) among medically admitted older adults [1], and reaches 50–80% in mechanically ventilated intensive care patients [2]. It independently raises in-hospital mortality, lengthens admission, and carries a near nine-fold risk of incident dementia [2,3].

What to prescribe, and whether to prescribe at all, is far less settled than daily practice suggests. Nikooie and colleagues found no difference in delirium severity between haloperidol and second-generation antipsychotics [4], and a 2025 meta-analysis of 11 randomised trials (n = 1,450) reported a standardised mean difference of −0.03 (95% CI −0.20 to 0.15) between classes, with greater extrapyramidal risk under haloperidol (odds ratio 2.72) [5]. The only adequately powered placebo-controlled trial in this field found that neither haloperidol nor risperidone outperformed placebo for delirium symptoms [6]. Multicomponent non-pharmacological programmes, by contrast, reduce incidence by roughly 43% and remain the evidence-based first line [7,8].

What clinicians actually prescribe, however, has diverged from that evidence and also differs sharply between health systems. In a nationwide United States cohort of 17,115 adults aged 65 years or older treated with an oral antipsychotic after major surgery between 2009 and 2018, quetiapine accounted for 53.0% of prescriptions and haloperidol use had declined over the study period, with no significant difference between agents in in-hospital death, arrhythmia, pneumonia or stroke [9]. An Irish national audit of 893 hospitalised patients with dementia found that 40% were started on a new or increased psychotropic during admission, frequently for delirium, and that documentation of risk and benefit, of discussion with patients and families, and of review for efficacy and tolerability was generally poor [10]. Two features recur across these audits: the choice of agent tracks local habit rather than trial evidence, and severity is rarely measured in a way that would allow the consequences of that choice to be examined.

Latin American data are thinner still. A prospective cohort at a university hospital in Bogotá screened 531 admissions and found delirium in 12%, rising to 31.8% on the geriatrics ward, with the mixed motor subtype predominating at 60.9% and age, four or more comorbidities and intensive care admission as the strongest associated factors [11]. Incidence studies of this kind establish the burden but do not describe what is prescribed once a liaison service is involved, nor how measured severity changes afterwards. We are not aware of a Latin American series that pairs prescribing patterns with serial severity scoring across an entire consultation-liaison caseload, so it remains unknown whether the haloperidol-dominant pattern of earlier decades persists here and what severity trajectories accompany it.

We therefore describe antipsychotic prescribing and serially measured delirium severity in consecutive inpatients seen by a consultation-liaison psychiatry service at a Mexican tertiary referral centre. The study is descriptive by design and by the limits of its data. Its contribution is a verified, fully reconciled account of real-world practice and measured outcomes from a region absent from this literature — not a comparison of drug efficacy, which this design cannot support. Reporting follows the STROBE statement for cohort studies [12].

2. Materials and Methods

2.1. Design and setting

We performed a retrospective, observational, longitudinal cohort study at the Consultation-Liaison Psychiatry service of Centro Médico Nacional "20 de Noviembre," ISSSTE, Mexico City, covering consultations between March 2021 and August 2022. The Research, Ethics and Biosecurity Committees of the institution approved the study on 2 October 2024 (protocol RPI 643.2024). The completed STROBE checklist accompanies this manuscript as Supplementary Material S1.

2.2. Participants and eligibility

We included adults aged 18 years or older who were seen in consultation for delirium diagnosed by the liaison psychiatry team according to DSM-5 criteria, supported by a positive CAM diagnostic algorithm [13], and who received at least one antipsychotic under psychiatric supervision. We excluded patients managed exclusively by geriatrics, patients whose delirium was confined to the intensive care unit without ward transfer, records without any documented CAM assessment, and patients under 18 years of age.

Three distinct constructs must be kept apart. The CAM diagnostic algorithm — acute onset and fluctuating course, inattention, and either disorganised thinking or altered level of consciousness — determines whether delirium is present. The DSM-5 criteria constitute the clinical diagnosis of record. The summed CAM-S short-form score (range 0–7) is a severity gradient, not a diagnostic threshold [14]. Eligibility rested on the first two; the summed CAM-S score was never an eligibility criterion. Consequently 14 patients (10.3%) had a first recorded CAM-S below 3, most often because the first psychiatric visit occurred after symptoms had already begun to settle or during a lucid interval of a fluctuating course. Section 3.3 reports this distribution as observed.

2.3. Measurements, assessors and schedule

Members of the consultation-liaison psychiatry team — attending liaison psychiatrists and senior psychiatry residents supervised by them — performed all CAM assessments at the bedside during routine care, using the instrument as applied in standard departmental practice. Assessments were scheduled at each psychiatric visit rather than at fixed calendar intervals, which is why the number of assessments per patient varied (median 5, IQR 2–7, range 1–22) and why follow-up ended when the liaison consultation ended, whether through discharge, transfer, death, or clinical resolution. No study-specific assessment was performed, no formal inter-rater reliability study was conducted during the care period, and assessors were not blinded to treatment. These are limitations of a records-based design and we do not present them otherwise.

2.4. Data verification and unit of analysis

Before any analysis we reconciled the study database against the extraction log under rules fixed in advance. One episode had been captured twice under identical record, admission date and length of stay; we retained the more complete of the two entries. One patient contributed two admissions roughly one year apart; we retained the first episode so that observations remained independent. Two further pairs of records shared an identical clinical-record code but belonged to demonstrably different patients — differing in name, sex, age, treating service and admission date — and were retained as separate patients, the shared code being a transcription error. One patient aged 14 years had entered the database in breach of the age criterion and was excluded at this stage. One CAM-S entry of 9, impossible on a 0–7 scale and

embedded in a sequence of scores of 1 followed by 0, was set to missing. Figure 1 traces every one of these steps.

Each patient contributes exactly one index antipsychotic regimen. The three regimens combining an antipsychotic with a sedative (dexmedetomidine in one patient, lorazepam in two) pair an antipsychotic with a non-antipsychotic, and we classify each by its index antipsychotic, with adjunctive sedative use retained as a covariate. Patients and exposures are therefore identical in number throughout, and every denominator in this report derives from the single verified dataset described here.

2.5. Outcomes

We defined outcomes as follows. Severity was the summed CAM-S score at each assessment, analysed longitudinally. Resolution was the first assessment at which CAM-S reached 0, with time measured from the first psychiatric assessment; a patient who later relapsed still counted as having resolved at that first occurrence, and this first-event definition is what the cumulative incidence curves estimate. Sustained severity control was a mean CAM-S of 2 or less across all of a patient's assessments; we retain this measure for continuity with the earlier report while noting that the threshold was chosen by the study team and has no external validation. Documented discharge improvement was the treating physician's global judgement recorded in the chart, which involves no instrument and no CAM threshold. Death, discharge without resolution, and transfer were treated explicitly in the time-to-event analysis rather than pooled.

2.6. Statistical analysis

We analysed the data in Python 3.14 using pandas, scipy, statsmodels and lifelines; the analysis script that generated every number, table and figure in this report accompanies the submission as Supplementary Material S2.

We summarised continuous variables as mean ± standard deviation with median and interquartile range, and categorical variables as counts with percentages, stating the denominator wherever it departs from the full cohort. Missing data were not imputed. Age was missing for one patient, length of stay and discharge disposition for one patient, and weight or height for ten patients, whose body-mass index is therefore reported on 126 valid observations. CAM-S assessments were missing by design rather than at random: follow-up ended when the psychiatric consultation ended, and patients who improved were discharged and stopped being assessed. This mechanism is informative, and we treat it as such throughout rather than assuming data missing at random.

We modelled the CAM-S trajectory with linear mixed-effects models including a random intercept and a random slope for time within patient, with time measured in days from each patient's first assessment. The primary trajectory model estimated the mean rate of change across the cohort. A second model added index agent and baseline CAM-S as fixed effects to estimate agent-associated differences, and a third added an agent-by-time interaction. We report coefficients with 95% confidence intervals throughout. Quetiapine (n = 2) was excluded from agent models, which cannot support a two-patient group.

For time to first resolution we estimated cumulative incidence with the Aalen–Johansen estimator, treating in-hospital death before resolution as a competing event and end of psychiatric follow-up without resolution as censoring; for patients who died, the competing event was placed at the last recorded

assessment, since the date of death was not available in the extraction. We report the Kaplan–Meier estimate alongside it to quantify the bias introduced by treating deaths as non-informative censoring, and we report numbers at risk, resolutions, deaths and censorings at each displayed time point so that the curves can be reconstructed. Baseline predictors of resolution were examined in a cause-specific Cox model restricted to four variables available before outcome assessment — baseline CAM-S, age, chronic kidney disease and sex — chosen in advance to respect the ratio of 35 events to covariates. We assessed proportional hazards with scaled Schoenfeld residuals and corrected discrimination for optimism using 500 bootstrap resamples. Length of stay, which the earlier report treated as a predictor, accrues during the same admission and can be a consequence of persistent delirium; we removed it, and with it the quasi-complete separation it produced.

This study tests no prespecified primary hypothesis. All agent comparisons, comorbidity analyses and subgroup estimates are exploratory. We therefore emphasise estimates and intervals over significance tests, and where we report the eight comorbidity models we note that no result survives or requires Holm correction because none reaches nominal significance. We set α at 0.05 for descriptive reference only. All estimates are associations; nothing here is a causal estimate, and no propensity-score or instrumental-variable method could repair the absence of dose, timing, motor subtype and non-pharmacological co-intervention data.

2.7. Ethics, record access and confidentiality

Care occurred between March 2021 and August 2022. The institutional committees approved the protocol on 2 October 2024, and research access to clinical records began only after that approval and the accompanying waiver of informed consent. Record retrieval and data extraction from the SIAH electronic record system ran from October 2024 to January 2025; de-identification was completed at extraction, before analysis, and all analyses reported here were performed on the de-identified file. The committees classified the study as "sin riesgo" under the Reglamento de la Ley General de Salud en Materia de Investigación para la Salud (Título II, Capítulo I, Artículo 17) and waived informed consent given the retrospective, records-based, standard-care-only design. The study conformed to the Declaration of Helsinki.

3. Results

3.1. Cohort and baseline characteristics

We screened 195 consultation-liaison records and excluded 57 (3 aged under 18 years, 31 without a documented CAM assessment, 23 with incomplete records), leaving 138 eligible records. These were entered into the study database as 139 rows, because one episode had been captured twice. Removing that duplicate capture left 138 delirium episodes; retaining only the first episode of one patient readmitted about a year later left 137 unique patients; and excluding one patient aged under 18 years, detected at verification, left an analytic cohort of 136 patients who contributed 724 CAM-S assessments (Figure 1).

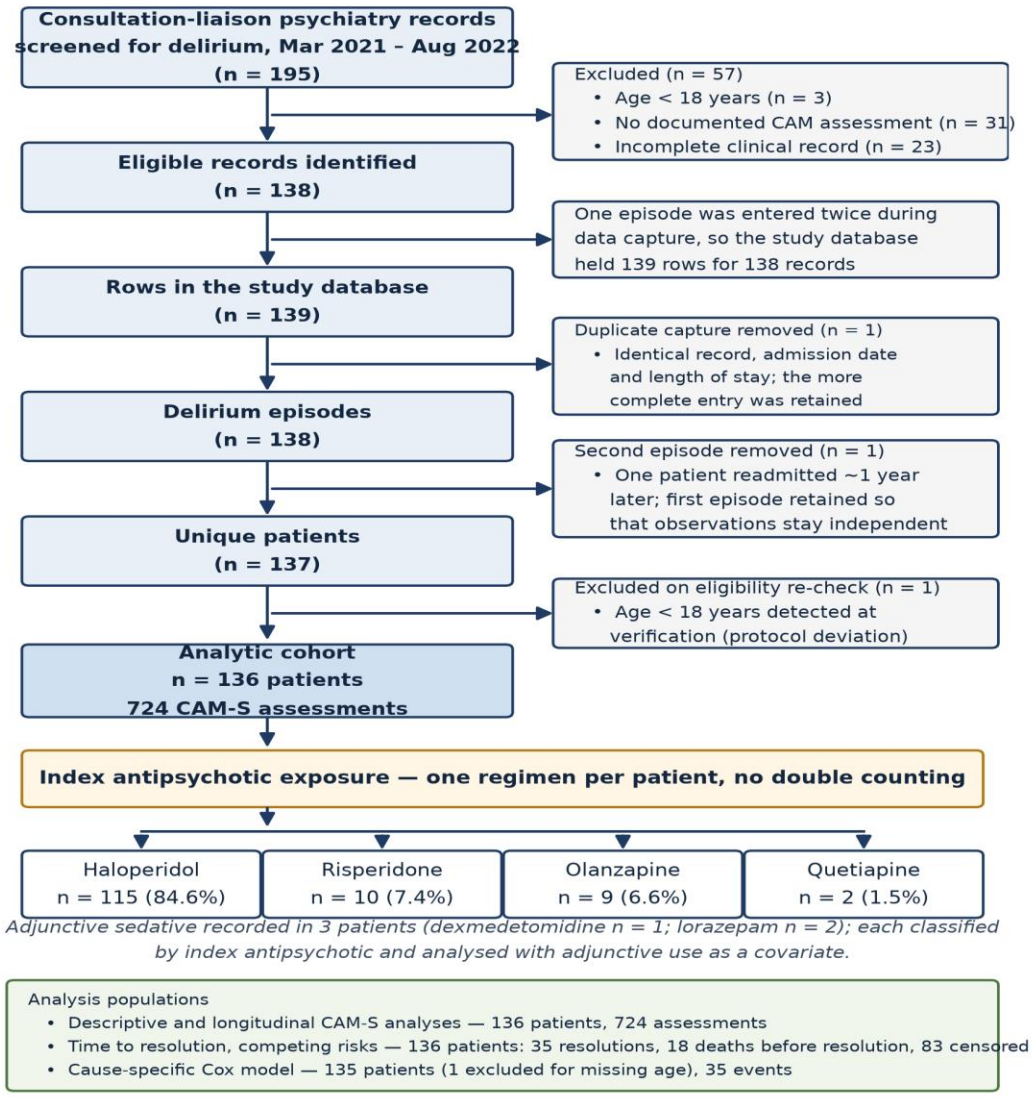


Figure 1. Patient flow from initial identification to the analytic cohort. Of 195 screened records, 57 were excluded, leaving 138 eligible records; these were held in the study database as 139 rows because one episode was captured twice. Successive removal of that duplicate capture (139 → 138 episodes), of the second episode of one readmitted patient (138 → 137 unique patients) and of one patient aged under 18 years detected at verification (137 → 136) yielded the analytic cohort. Every count in this manuscript derives from this account. The three regimens pairing an antipsychotic with a sedative are classified by their index antipsychotic; no patient contributes to more than one agent group.

Eighty-four patients were men (61.8%). Mean age was 67.1 ± 13.1 years (median 68, IQR 58–77, range 29–91; missing in one patient). Mean body-mass index was 26.8 ± 4.8 kg/m² among the 126 patients with valid measurements, of whom 79 (62.7%) were overweight or obese. Recent surgery was the most common precipitating context (99 patients, 72.8%), followed by hypertension (85, 62.5%), type 2 diabetes (61, 44.9%), active infection (50, 36.8%), prior delirium (34, 25.0%), heart failure (33, 24.3%), chronic kidney disease (29, 21.3%), acute kidney injury (23, 16.9%) and dementia (5, 3.7%). Median length of stay was 16 days (IQR 10–26, range 1–90). Median psychiatric follow-up was 5 days (IQR 3–8, range 1–22). Mean CAM-S at the first assessment was 3.79 ± 1.27 (median 4, IQR 3–4); 14 patients (10.3%) scored below 3, for the reasons set out

in Section 2.2. The first assessment fell on the first day of the recorded window for 99 patients and later for 37. Tables 1 and 2 and Figures 2–4 present these characteristics.

Table 1. Baseline demographic and clinical characteristics of the analytic cohort (n = 136).

Characteristic	n (%) or mean ± SD	Median (IQR); range
Male sex	84 (61.8%)	—
Age, years (n = 135)	67.1 ± 13.1	68 (58–77); 29–91
Body-mass index, kg/m² (n = 126)	26.8 ± 4.8	26.2 (23.5–29.7); 15.2–44.4
Overweight or obesity (BMI ≥ 25; n = 126)	79 (62.7%)	—
Recent surgery	99 (72.8%)	—
Hypertension	85 (62.5%)	—
Type 2 diabetes mellitus	61 (44.9%)	—
Active infection	50 (36.8%)	—
Prior delirium	34 (25.0%)	—
Heart failure	33 (24.3%)	—
Chronic kidney disease	29 (21.3%)	—
Acute kidney injury	23 (16.9%)	—
Dementia	5 (3.7%)	—
Psychiatric follow-up, days	5.8 ± 4.5	5 (3–8); 1–22
Length of stay, days (n = 135)	19.8 ± 14.3	16 (10–26); 1–90
CAM-S at first assessment	3.79 ± 1.27	4 (3–4); 0–7
CAM-S at last assessment	2.21 ± 1.77	2 (1–4); 0–7
Assessments per patient	5.3 ± 4.2	5 (2–7); 1–22

Denominators are stated wherever they depart from 136. Age and length of stay were missing for one patient each; body-mass index could not be computed for ten patients whose recorded weight or height was zero. CAM-S is the summed short-form Confusion Assessment Method severity score (range 0–7).

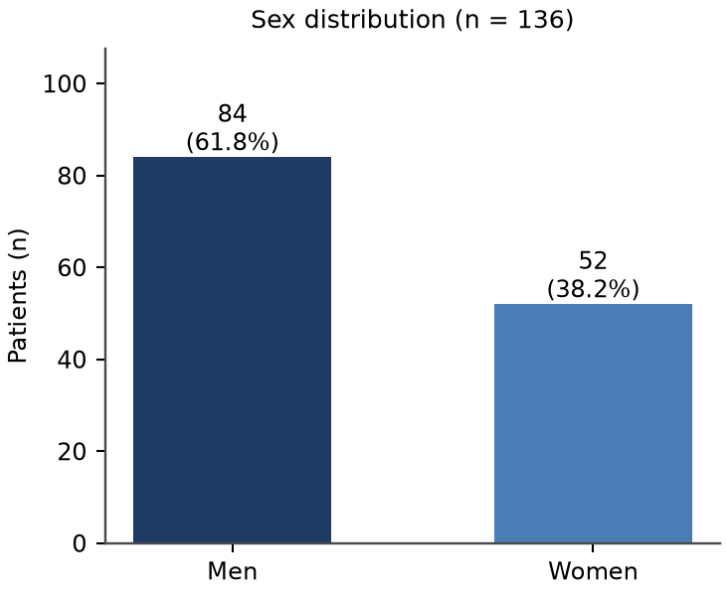


Figure 2. Sex distribution of the analytic cohort.

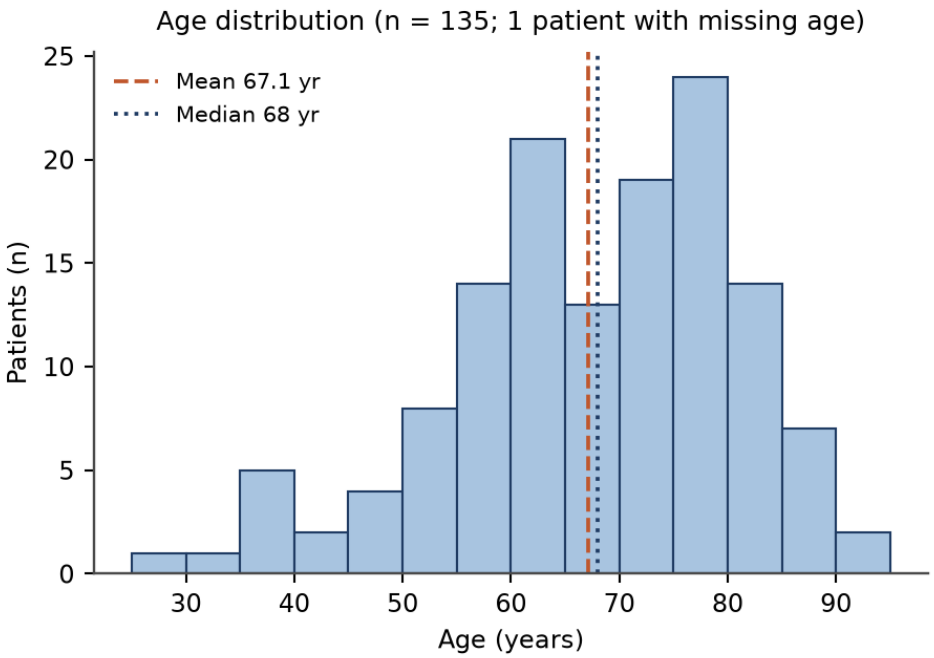


Figure 3. Age distribution. Age was missing for one patient, who is excluded from this figure and from age-adjusted models.

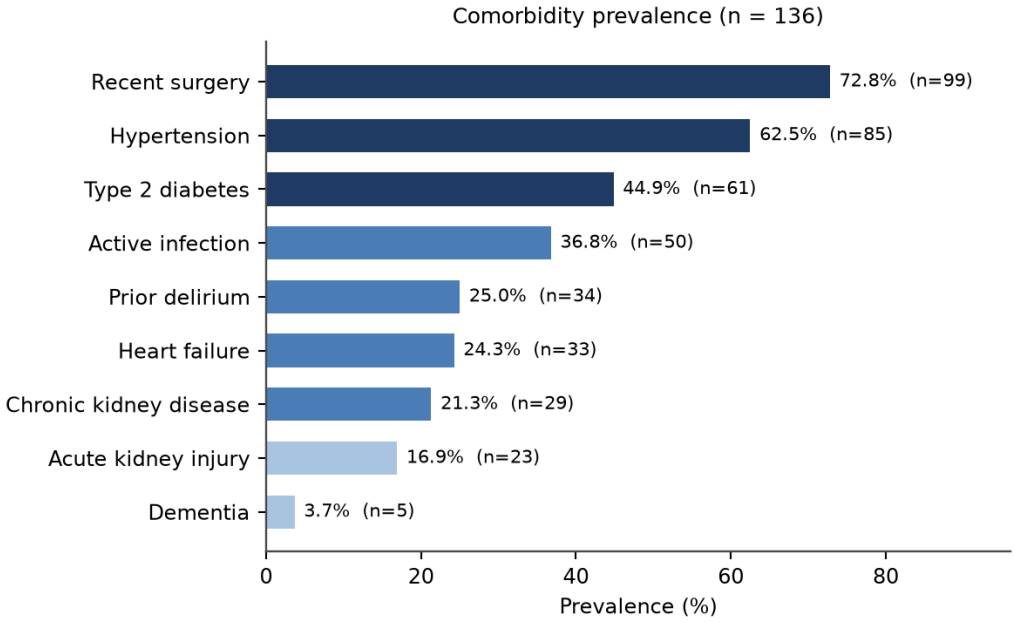


Figure 4. Prevalence of recorded comorbidities and precipitating context. Denominator is the full cohort of 136 patients for every bar.

3.2. Antipsychotic prescribing

Haloperidol accounted for 115 of 136 regimens (84.6%), risperidone for 10 (7.4%), olanzapine for 9 (6.6%) and quetiapine for 2 (1.5%). Three patients received an adjunctive sedative. Mean CAM-S per patient across all assessments was 2.99 under haloperidol, 3.20 under risperidone, 1.95 under olanzapine and 2.80 under quetiapine (Table 3, Figure 5). Baseline severity differed across these groups in ways that track prescribing

choice: mean baseline CAM-S was 3.76 ± 1.25 under haloperidol, 4.10 ± 1.85 under risperidone, 3.56 ± 0.53 under olanzapine and 5.00 ± 1.41 under quetiapine. With nine olanzapine-treated patients and two receiving quetiapine, these group means carry wide and overlapping intervals, and Figure 5 displays them accordingly, and Figure 6 shows the corresponding severity-control outcomes with their Wilson confidence intervals.

Table 2. Baseline characteristics by index antipsychotic.

Characteristic	Haloperidol (n = 115)	Risperidone (n = 10)	Olanzapine (n = 9)	Quetiapine (n = 2)
Male sex, n (%)	69 (60.0%)	7 (70.0%)	6 (66.7%)	2 (100.0%)
Age, years, mean ± SD	67.2 ± 12.2	67.0 ± 15.7	61.8 ± 18.2	89.5 ± 2.1
Body-mass index, mean ± SD	27.1 ± 4.9	24.2 ± 4.0	26.2 ± 4.8	24.3 ± 3.5
Hypertension, n (%)	72 (62.6%)	7 (70.0%)	4 (44.4%)	2 (100.0%)
Type 2 diabetes, n (%)	54 (47.0%)	3 (30.0%)	3 (33.3%)	1 (50.0%)
Chronic kidney disease, n (%)	25 (21.7%)	4 (40.0%)	0 (0.0%)	0 (0.0%)
Acute kidney injury, n (%)	18 (15.7%)	4 (40.0%)	1 (11.1%)	0 (0.0%)
Heart failure, n (%)	29 (25.2%)	3 (30.0%)	0 (0.0%)	1 (50.0%)
Active infection, n (%)	44 (38.3%)	4 (40.0%)	2 (22.2%)	0 (0.0%)
Recent surgery, n (%)	83 (72.2%)	7 (70.0%)	8 (88.9%)	1 (50.0%)
Prior delirium, n (%)	28 (24.3%)	4 (40.0%)	1 (11.1%)	1 (50.0%)
CAM-S at first assessment, mean ± SD	3.76 ± 1.25	4.10 ± 1.85	3.56 ± 0.53	5.00 ± 1.41
Mean CAM-S across assessments	2.99	3.20	1.95	2.80
Length of stay, days, median (IQR)	16 (9–25)	22 (11–26)	12 (9–17)	27 (24–30)
Adjunctive sedative, n	1	1	0	1

Groups differ at baseline in ways that follow prescribing choice rather than randomisation. No between-group significance tests are presented: with nine olanzapine-treated and two quetiapine-treated patients, such tests would be uninformative, and the imbalance is not a hypothesis to be tested but a reason not to compare agents. Percentages are within-column.

Table 3. Prescribing frequency and observed CAM-S outcomes by index antipsychotic.

Index antipsychotic	Patients, n (%)	Mean CAM-S per patient (95% CI)	CAM-S at first assessment, mean ± SD	Reached CAM-S = 0, n (%)	Mean CAM-S ≤ 2, n (%)
Haloperidol	115 (84.6%)	2.99 (2.74–3.24)	3.76 ± 1.25	29 (25.2%)	27 (23.5%)
Risperidone	10 (7.4%)	3.20 (2.17–4.23)	4.10 ± 1.85	2 (20.0%)	2 (20.0%)
Olanzapine	9 (6.6%)	1.95 (1.38–2.52)	3.56 ± 0.53	4 (44.4%)	5 (55.6%)
Quetiapine	2 (1.5%)	2.80 (not estimable)	5.00 ± 1.41	0 (0.0%)	0 (0.0%)
All patients	**136 (100%)**	**2.93 (2.71–3.16)**	**3.79 ± 1.27**	**35 (25.7%)**	**34 (25.0%)**

Confidence intervals are for the group mean and are not adjusted for multiplicity or for baseline severity. Ranking agents by these means would misrepresent the data; the intervals overlap and the groups differ systematically at baseline (Table 2).

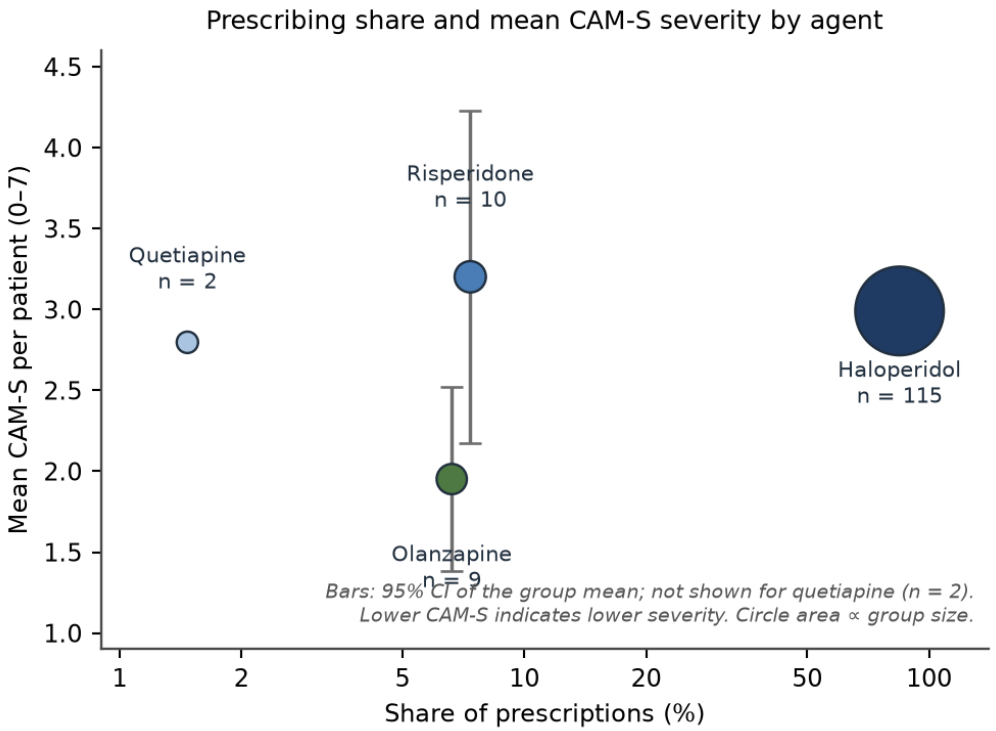


Figure 5. Share of prescriptions and mean CAM-S per patient, by index antipsychotic. Error bars are 95% confidence intervals of the group mean and are omitted for quetiapine (n = 2), for which no interpretable interval exists. Circle area is proportional to group size. The intervals overlap across agents; the figure describes prescribing and observed severity, not comparative effectiveness.

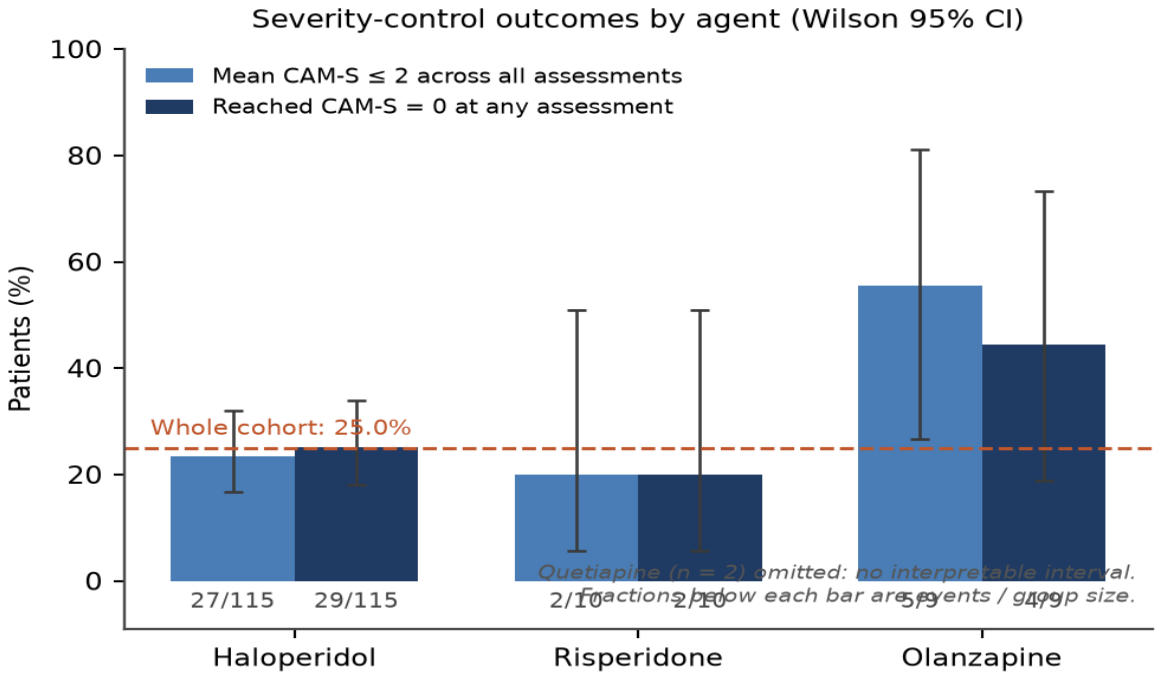


Figure 6. Severity-control outcomes by agent, with Wilson 95% confidence intervals. Fractions below each bar give events over group size. Quetiapine (n = 2) is omitted. The width of the risperidone and olanzapine intervals is the substantive finding: these groups cannot discriminate between agents.

3.3. Severity over time

CAM-S declined by 0.38 points per day (95% CI 0.30–0.45; $p < 0.001$) in the mixed-effects model, from a cohort mean of 3.79 at the first assessment to 2.21 at the last. The mean within-patient reduction was 1.57 points (95% CI 1.28–1.87; Wilcoxon signed-rank $p < 0.001$). Figure 7 shows the distribution of summed CAM-S at the first and the last assessment, including the 14 patients whose first recorded score fell below 3.

Figure 8 shows the observed data rather than a smoothed curve, together with the number of assessments contributing to each day. The mean falls steeply through day 4 and then flattens and rises. That rise is not a clinical deterioration of the cohort: it is attrition working in one direction. Patients whose delirium resolved were discharged and stopped being assessed, so the patients still under psychiatric follow-up after the first week are increasingly those who did not improve. Only 31 assessments contribute at day 7 and 7 at day 14. Any estimate drawn from the later part of this curve describes a shrinking and progressively more severe subgroup, and we make no claim from it.

Adjusted for baseline severity, no agent differed significantly from haloperidol in mean CAM-S: olanzapine −0.50 (95% CI −1.04 to 0.05; $p = 0.075$) and risperidone −0.23 (95% CI −0.69 to 0.24; $p = 0.34$). Adding an agent-by-time interaction produced a steeper decline under olanzapine (−0.45 points per day beyond the common slope; 95% CI −0.78 to −0.13; $p = 0.007$) and a shallower one under risperidone (+0.27; 95% CI 0.02 to 0.53; $p = 0.037$). These interaction estimates rest on nine and ten patients respectively, were not prespecified, and are reported as exploratory observations that would need replication in a design capable of testing them.

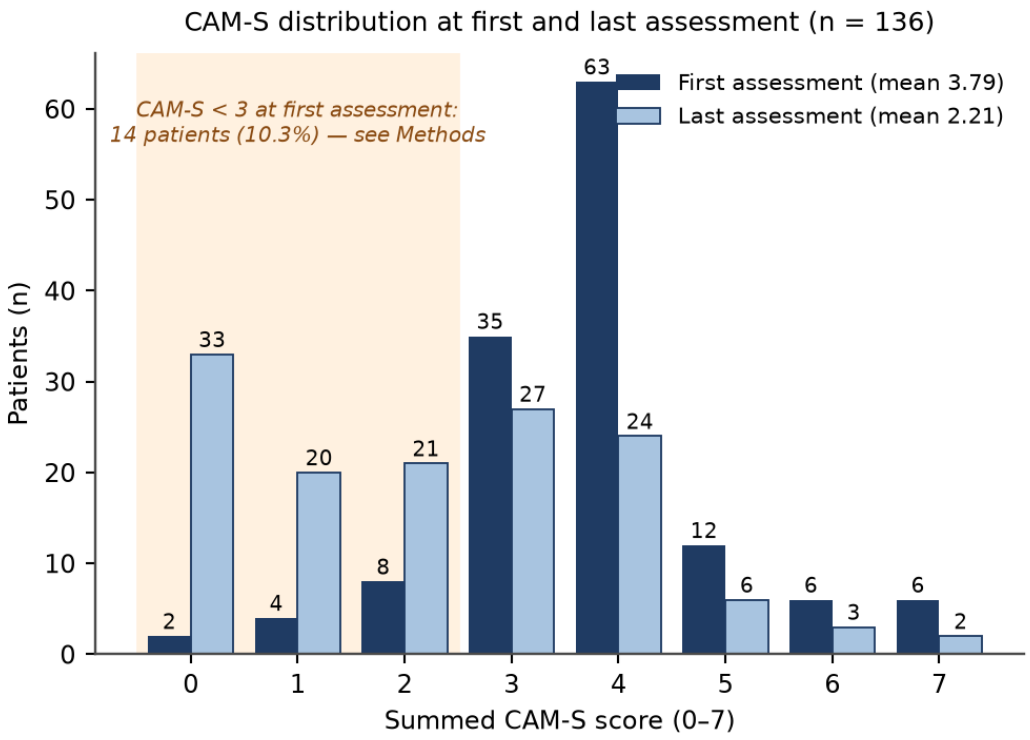


Figure 7. Distribution of summed CAM-S at the first and the last assessment. The shaded region marks scores below 3 at the first assessment, recorded in 14 patients (10.3%). Eligibility rested on the CAM diagnostic algorithm and DSM-5 criteria, not on the summed severity score; Section 2.2 explains the distinction.

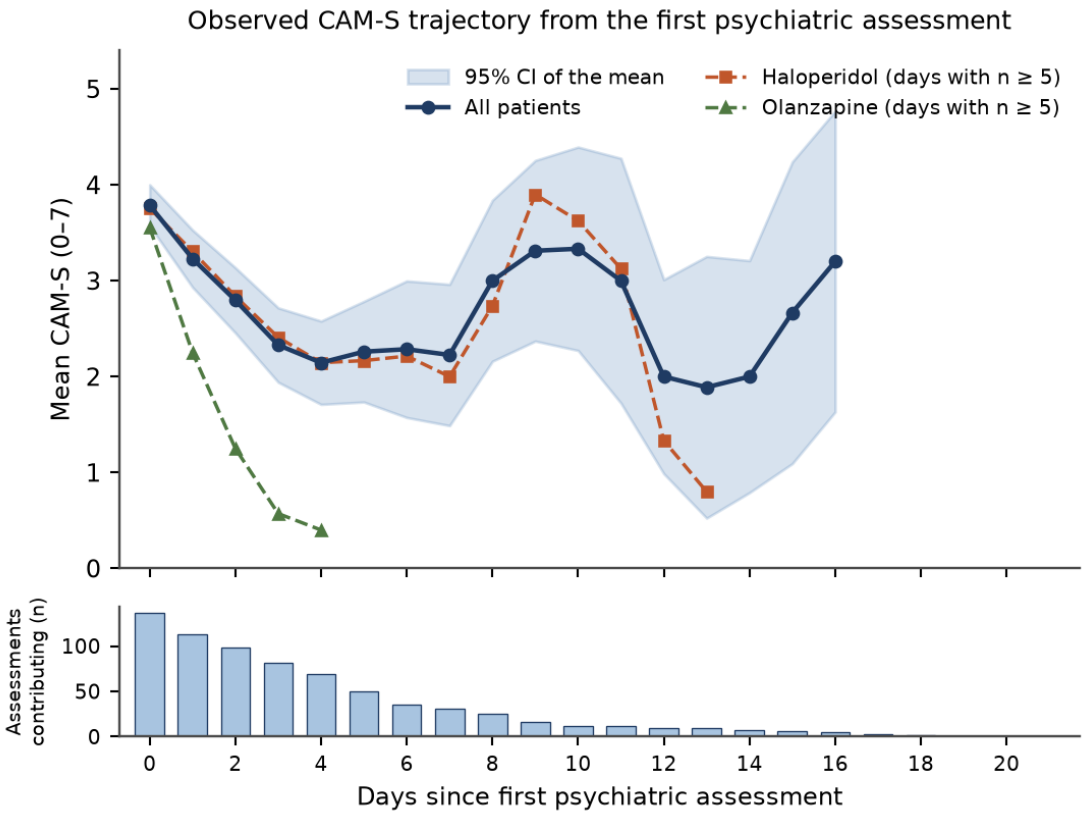


Figure 8. Observed mean CAM-S by day since the first psychiatric assessment, with the 95% confidence interval of the mean and the number of assessments contributing to each day. Only days with at least five assessments are plotted. The rise after day 5 reflects informative attrition rather than clinical deterioration: patients who improve are discharged and stop being assessed, so the remaining sample is progressively more severe.

3.4. Time to resolution

Thirty-five patients (25.7%) reached CAM-S = 0 at some point during follow-up; 18 died before resolution; 83 were censored when psychiatric follow-up ended without resolution. Among those who resolved, the median time from the first assessment was 3 days (IQR 2–4).

The cumulative incidence of first resolution was 13.8% (95% CI 7.8–21.5) by day 3, 32.1% (95% CI 22.4–42.3) by day 5, 38.3% (95% CI 27.1–49.4) by day 7 and 46.1% (95% CI 31.8–59.2) by day 14. The competing risk of death reached 17.2% by day 7 and 27.4% by day 14. The Kaplan–Meier estimator, which treats those deaths as non-informative censoring, gave 41.6% and 53.6% at the same time points, overstating resolution by 3.3 and 7.5 percentage points. Figure 9 reports numbers at risk, resolutions, deaths and censorings at each displayed day; 22 patients remained at risk at day 7, 12 at day 10 and 6 at day 14, which is why the confidence band widens as it does and why we quote no point estimate beyond day 14 without it.

These figures reconcile with the 33 patients (24.3%) whose final recorded CAM-S was 0. The two quantities differ by two patients who reached CAM-S = 0 and then scored above 0 at a later assessment. The cumulative incidence estimates first resolution; the cross-sectional count describes status at the last assessment; and neither is a rate of sustained recovery.

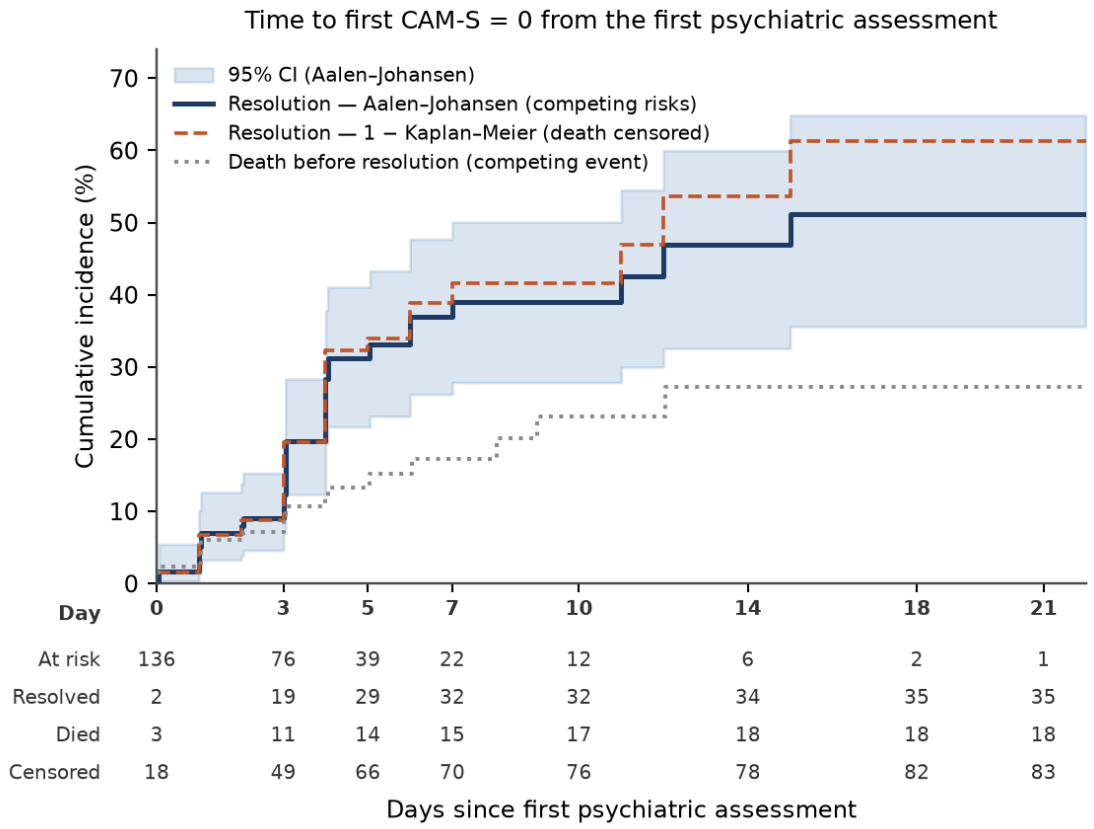


Figure 9. Cumulative incidence of first delirium resolution (CAM-S = 0) with in-hospital death as a competing event. The Aalen–Johansen estimate is the primary quantity; the Kaplan–Meier curve is shown to quantify the bias introduced by treating the 18 deaths as non-informative censoring. Counts below the panel give patients at risk, resolutions, deaths and censorings at each displayed day. Estimates beyond day 10 rest on 12 or fewer patients at risk.

3.5. Baseline predictors of resolution

In the cause-specific Cox model (135 patients, 35 events), baseline severity was the only variable associated with resolution: each additional CAM-S point at the first assessment corresponded to a hazard ratio of 0.49 (95% CI 0.36–0.67; $p < 0.001$). Age carried a hazard ratio of 0.78 per decade (95% CI 0.59–1.04; $p = 0.096$), chronic kidney disease 0.78 (95% CI 0.29–2.09; $p = 0.618$) and male sex 0.68 (95% CI 0.34–1.35; $p = 0.274$) (Table 4). The proportional-hazards assumption held. Discrimination was 0.708 apparent and 0.687 after bootstrap correction for optimism, and observed resolution by day 7 rose monotonically across model-defined risk tertiles (18.0%, 39.8% and 100%). Figure 10 presents the model.

We did not model persistent delirium as a binary outcome. Such a model would lack a fixed assessment time and a stable denominator, and length of stay greater than 20 days — the variable that would otherwise carry it — accrues during the same admission and produces quasi-complete separation. Neither age over 75 years nor chronic kidney disease survives as a predictor of resolution in this analysis. In an exploratory univariable analysis, prior delirium was associated with lower probability of resolution (hazard ratio 0.18; 95% CI 0.06–0.53; $p = 0.002$); with 35 events in total we did not add it to the multivariable model, and we flag it as a hypothesis rather than a finding.

Index agent entered the model as an exploratory covariate: olanzapine 2.80 (95% CI 0.91–8.57; p = 0.072) and risperidone 0.39 (95% CI 0.09–1.75; p = 0.221) relative to haloperidol. Nine and ten patients cannot resolve this question, and these intervals say so.

Table 4. Cause-specific Cox model for time to first delirium resolution (135 patients, 35 events).

Baseline variable	Hazard ratio	95% CI	p
CAM-S at first assessment (per point)	0.49	0.36 to 0.67	< 0.001
Age (per 10 years)	0.78	0.59 to 1.04	0.096
Chronic kidney disease	0.78	0.29 to 2.09	0.618
Male sex	0.68	0.34 to 1.35	0.274

Death before resolution was censored in this cause-specific model; the cumulative incidence in Figure 9 accounts for it as a competing event. Proportional hazards held on scaled Schoenfeld residuals. Apparent c-index 0.708; optimism 0.021 over 500 bootstrap resamples; optimism-corrected c-index 0.687.

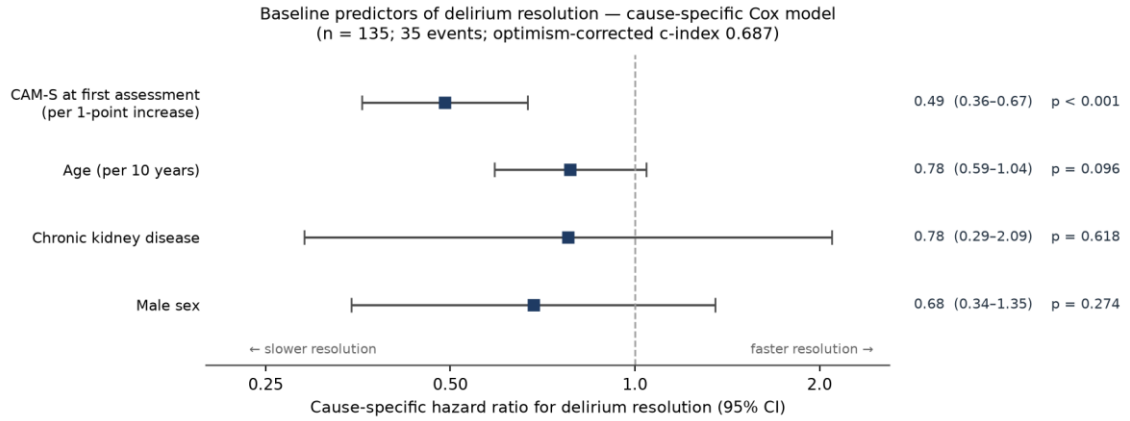


Figure 10. Cause-specific hazard ratios for delirium resolution from the multivariable Cox model, restricted to variables available before outcome assessment. Length of stay was not eligible as a baseline predictor because it accrues during the same admission and may follow from persistent delirium rather than precede it.

3.6. Comorbidity and severity trajectory

No comorbidity was associated with the CAM-S trajectory once within-patient correlation was modelled. Adjusted for time and baseline severity, coefficients were 0.005 for type 2 diabetes (95% CI −0.25 to 0.26; p = 0.97), −0.20 for hypertension (95% CI −0.46 to 0.06; p = 0.14), 0.24 for chronic kidney disease (95% CI −0.08 to 0.56; p = 0.14), 0.08 for acute kidney injury (95% CI −0.26 to 0.42; p = 0.65), 0.03 for heart failure (95% CI −0.25 to 0.32; p = 0.82), 0.17 for prior delirium (95% CI −0.14 to 0.47; p = 0.29), 0.16 for active infection (95% CI −0.11 to 0.42; p = 0.24) and −0.02 for recent surgery (95% CI −0.31 to 0.26; p = 0.88). Table 5 reports these estimates.

Table 5. Comorbidity and CAM-S trajectory: linear mixed-effects models adjusted for time and baseline severity.

Comorbidity	Coefficient (CAM-S points)	95% CI	p
Type 2 diabetes mellitus	0.005	−0.253 to 0.264	0.97
Hypertension	−0.199	−0.461 to 0.063	0.14
Chronic kidney disease	0.241	−0.078 to 0.560	0.14
Acute kidney injury	0.079	−0.259 to 0.417	0.65
Heart failure	0.034	−0.254 to 0.322	0.82

Prior delirium	0.166	−0.143 to 0.474	0.29
Active infection	0.157	−0.107 to 0.421	0.24
Recent surgery	−0.022	−0.308 to 0.264	0.88

Each row is a separate model with a random intercept and random slope for time within patient. Coefficients are differences in CAM-S points associated with the comorbidity. No estimate reaches nominal significance, so no multiplicity correction is required across these eight exploratory models.

An analysis that treated the repeated assessments from each patient as independent observations would show diabetes, chronic kidney disease, acute kidney injury, infection, surgery and prior delirium each retarding severity reduction, with p values as low as 0.0001. That approach inflates the effective sample size roughly five-fold and drives p values towards zero. Once the repeated-measures structure is modelled correctly, none of these associations remains.

3.7. Discharge outcomes and two different constructs

The treating physician documented clinical improvement at discharge in 106 of 136 patients (77.9%). Twenty patients died in hospital (14.7%), three left against medical advice (2.2%), three were discharged for other reasons (2.2%), two were transferred (1.5%) and one was discharged on maximum-benefit grounds (0.7%); disposition was missing for one patient (Figure 11).

Sustained severity control, defined as a mean CAM-S of 2 or less across all assessments, was reached by 34 patients (25.0%). The distance between 77.9% and 25.0% is not evidence about drug efficacy. One number is a physician's global judgement at a single moment, made in the context of a resolving medical illness; the other is an arbitrary threshold applied to a severity score averaged over a variable number of assessments, and averaging penalises patients assessed often and early. We report both because both appear in the record, and we draw no inference about pharmacological benefit from their difference. The comparison a reader might want — measured delirium outcomes with and without antipsychotic treatment — cannot be made here, because every patient in this cohort received an antipsychotic.

Among the 126 patients with valid anthropometry, mean CAM-S was highest in the underweight (3.7) and obesity grade II–III strata (3.5 and 4.0), and lowest in the normal-weight and overweight strata (2.9 and 2.8); the confidence intervals overlap throughout and the extreme strata contain 2 and 1 patients (Figure 12). We report this pattern as description only.

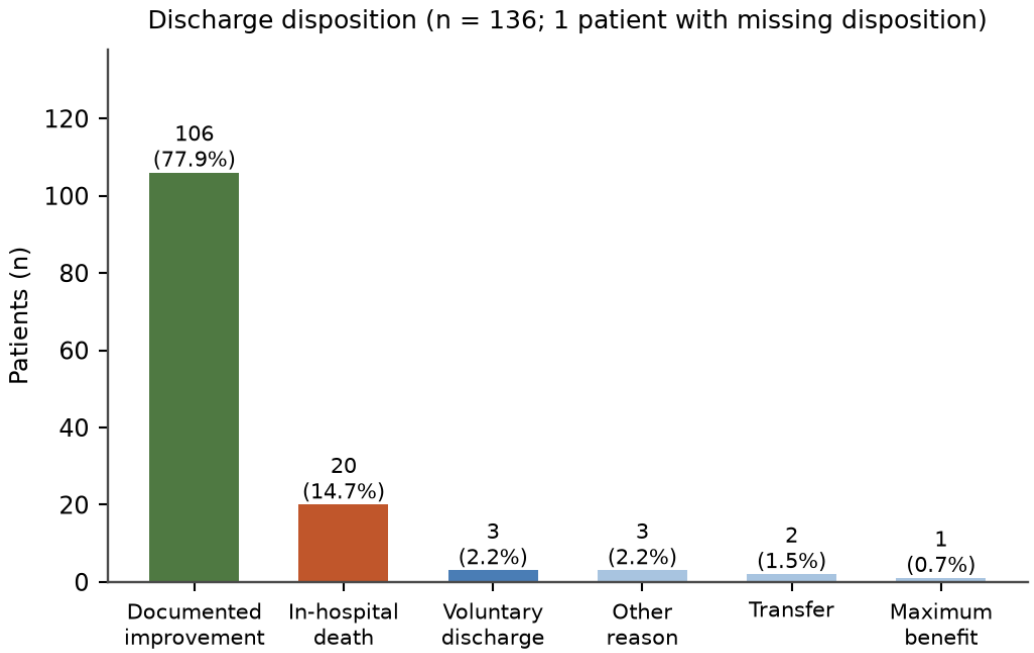


Figure 11. Discharge disposition. Disposition was missing for one patient, who is not represented in any bar.

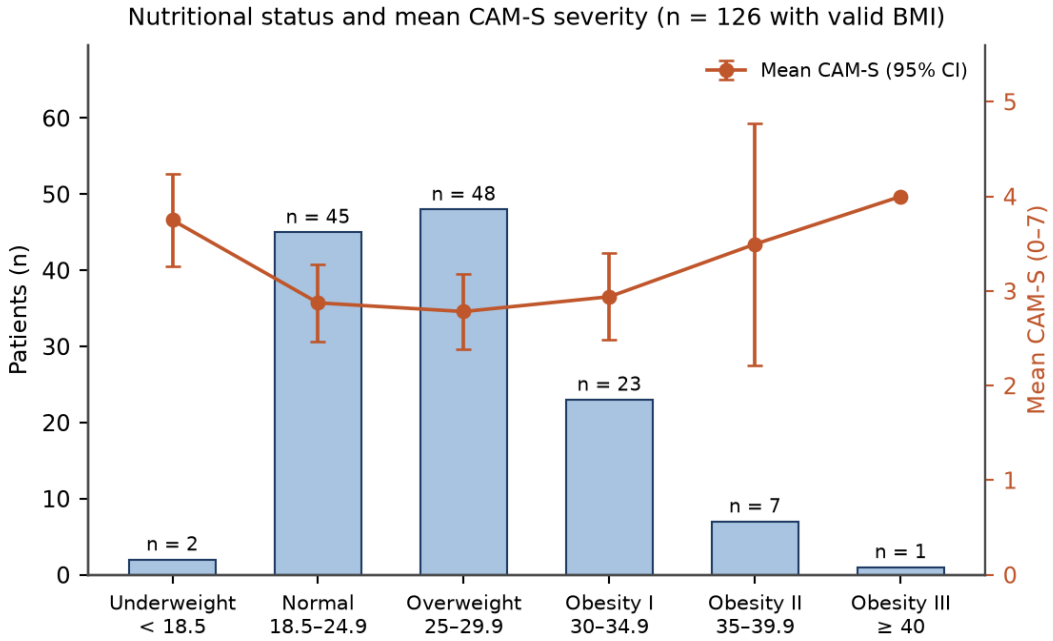


Figure 12. Nutritional status and mean CAM-S among the 126 patients with valid anthropometry, with 95% confidence intervals of the stratum mean. The underweight and obesity grade III strata contain two patients and one patient respectively.

4. Discussion

Antipsychotic prescribing for delirium at this centre was concentrated almost entirely on haloperidol, which accounted for 115 of 136 regimens (84.6%). Measured severity fell steadily under that practice, by 0.38 CAM-S points per day (95% CI 0.30–0.45), and 38.3% of patients (95% CI 27.1–49.4) had reached a score

of zero by day 7. Adjusted for baseline severity, no agent differed significantly from haloperidol. The one baseline variable that tracked resolution was how severe the delirium was when psychiatry first saw the patient.

That degree of concentration is worth placing against practice elsewhere, because the two diverge. In the United States nationwide cohort of older adults treated after major surgery, quetiapine accounted for 53.0% of oral antipsychotic prescriptions and haloperidol use had fallen over the study decade, with no difference between agents in in-hospital adverse events [9]. Our centre sits at the opposite end: haloperidol accounted for 84.6% of regimens, and the atypical agents together for 15.4%. Neither pattern follows from the trial evidence, which distinguishes between these agents on tolerability rather than efficacy. What differs between the two settings is which habit prevailed, and the Irish audit suggests why habit has room to prevail: psychotropics are frequently started during admission with little documented weighing of risk and benefit and little structured review of whether they worked [10]. Prescribing tradition, in other words, fills a space that evidence has left open, and it fills it differently in different systems. Nikooie and colleagues found no severity difference between haloperidol and second-generation agents [4]; the 2025 meta-analysis of 11 randomised trials found a standardised mean difference of −0.03 (95% CI −0.20 to 0.15) between classes [5]. The trials that matter most are placebo-controlled. Agar and colleagues found that neither haloperidol nor risperidone outperformed placebo, with more adverse events under risperidone [6]. HOPE-ICU found no effect of intravenous haloperidol on days alive without delirium or coma [15]. MIND-USA, with 566 patients, found no difference between haloperidol, ziprasidone and placebo in days alive without delirium or coma (haloperidol versus placebo: adjusted odds ratio 1.14, 95% CI 0.83–1.56; p = 0.41) or in 90-day mortality [16]. AID-ICU, the largest such trial with 1,000 patients, found no benefit of haloperidol on days alive and out of hospital at 90 days in its primary analysis, although a pre-planned Bayesian secondary analysis estimated a 92% posterior probability of some benefit and a signal towards lower 90-day mortality [17]. A phase III trial of olanzapine versus haloperidol in advanced cancer was stopped for futility [18]. Subsequent syntheses have not changed this picture: a 2024 meta-analysis of five ICU trials (n = 1,750) found no increase in delirium- or coma-free days [19]; a 2025 meta-analysis of haloperidol in ICU delirium (n = 2,863) found no reduction in delirium or mortality [20]; a 2023 meta-analysis of olanzapine in critically ill adults found no symptom relief, though less hypotension than comparators [21]; and a 2024 umbrella review identified dexmedetomidine and multicomponent non-pharmacological care, not antipsychotics, as carrying the strongest preventive evidence [22]. The documented advantage of atypical agents is tolerability, principally fewer extrapyramidal events, rather than efficacy [5,23,24]. Prescribing guidance for older adults continues to counsel caution with all antipsychotics [25], and reviews of contemporary practice describe the same haloperidol-dominant pattern we observed [26].

Our data cannot adjudicate between agents, and we do not present them as if they could. Nine patients received olanzapine and two received quetiapine. Dose, route, timing relative to symptom onset, switching, motor subtype and non-pharmacological co-interventions were not recorded, so the standard machinery for handling confounding by indication has nothing to work with; adjusting for baseline CAM-S, kidney disease and heart failure would address a small fraction of the problem while creating the impression that the rest had been handled. The steeper CAM-S decline associated with olanzapine in the interaction model is compatible with a drug effect, with selection of patients whose delirium was already resolving, and with chance across an unplanned comparison. We report it because it is in the data; we would not act on it.

The gap between physician-documented improvement at discharge (77.9%) and instrument-based sustained severity control (25.0%) invites over-interpretation. Reading that gap as evidence that antipsychotics add little beyond natural recovery would require a comparison this cohort does not contain, since every patient received an antipsychotic. The two numbers measure different things on different timescales: a clinician's global judgement at the moment of discharge, when the precipitating illness has usually resolved, against a threshold applied to an average of serial scores. Averaging is itself unkind to patients assessed frequently and early. The honest reading is narrower and still useful: charted improvement should not be treated as a proxy for measured delirium severity in this setting, and services that rely on it will overestimate how many of their patients have had their delirium controlled.

Three findings that a reader might expect from a cohort of this kind are not supported once the data are analysed correctly, and we state that plainly. First, the comorbidity effects: diabetes, kidney disease, infection, surgery and prior delirium appeared to retard recovery only because repeated assessments from the same patient were treated as independent. Modelling within-patient correlation removes every one of these associations. Second, the predictors of persistent delirium: age over 75 years and chronic kidney disease do not predict resolution in a properly specified model, and length of stay over 20 days was never a legitimate baseline predictor, since it accrues during the same admission and may follow from persistent delirium rather than cause it. Third, the reported superiority of olanzapine after covariate adjustment: the omnibus test was never significant, the adjusted comparison does not reach significance in the present analysis, and describing an effect as “confirmed” on the strength of a selected subgroup contrast was wrong.

Two points about representation of the evidence deserve correction. The 2026 Society of Critical Care Medicine guideline on caring for older adults in the ICU makes no recommendation regarding antipsychotic treatment of established delirium in older critically ill adults; its conditional recommendation against antipsychotics concerns prevention [27]. Separately, the 2018 PADIS guideline conditionally recommends against routine use of haloperidol or atypical antipsychotics to treat delirium in ICU patients [28], a recommendation that concerns intensive care rather than the general-ward and post-surgical population described here.

4.1. Strengths

The cohort contributes serial severity measurement, rather than a single assessment, from a region with almost no published pharmacoepidemiological data of this kind: 724 CAM-S assessments across 136 patients, collected during routine liaison care over 18 months. Measures that capture the intensity and duration of delirium, not merely its presence, carry the strongest associations with clinical outcomes [29]. Every number in this report derives from one verified dataset and one analysis script, with the patient-flow account reconciled from initial identification through each analysis population, and with the discrepancies found during verification reported rather than smoothed over. The time-to-event analysis handles death as a competing event and reports the numbers at risk that make the curves reconstructible.

4.2. Limitations

This is a single-centre retrospective study of records generated for clinical rather than research purposes, and its limits are substantial. Doses, routes, timing of initiation, switching and treatment duration were not recorded, so a patient given haloperidol 0.5 mg once and a patient given 10 mg daily appear as the same exposure. Motor subtype was not systematically documented. Non-pharmacological care was not recorded at all, which matters greatly given that multicomponent programmes carry the strongest evidence in this field [7]. No safety data were available: no QTc measurements, extrapyramidal symptoms, sedation, falls or metabolic parameters, so the tolerability differences that constitute the real distinction between these agents cannot be examined. There was no untreated comparison group. Assessors were not blinded, no inter-rater reliability study was conducted, and the CAM-S short form (0–7) is coarser than the long form (0–19) or the DRS-R-98. Follow-up ended with the psychiatric consultation, generating attrition that is informative rather than random and that we have not attempted to correct by modelling; estimates beyond the first week rest on fewer than a dozen patients. Body-mass index is missing for ten patients, age and disposition for one each. With 35 resolution events, the Cox model tolerates four covariates and no more, and its optimism-corrected c-index of 0.687 indicates modest discrimination that supports group-level description, not individual prediction. The mean CAM-S ≤ 2 threshold remains an internal convention without external validation. Finally, the agent groups outside haloperidol contain nine, ten and two patients; no analytic method converts that into a comparison of drugs.

5. Conclusions

Antipsychotic prescribing for delirium at this Mexican tertiary referral centre was concentrated on haloperidol, and measured delirium severity improved steadily under routine liaison care, with baseline severity the only baseline characteristic that tracked the probability of resolution. These data describe practice and outcomes at one centre; they do not identify a causal effect of any agent, and the agent groups are too small to support one.

Two implications follow for services that manage delirium in this setting. Documented improvement at discharge and instrument-measured severity control are not interchangeable, and services that audit only the former will substantially overestimate delirium control. Prospective cohorts in this region should record dose, timing, motor subtype, non-pharmacological co-interventions and safety endpoints, and should schedule assessments at fixed intervals independent of the consultation, so that the attrition that limits this analysis does not limit theirs.

List of Abbreviations

BMI, body mass index; CAM, Confusion Assessment Method; CAM-S, Confusion Assessment Method–Severity; CI, confidence interval; DRS-R-98, Delirium Rating Scale–Revised–98; DSM-5, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; ICU, intensive care unit; IQR, interquartile range; ISSSTE, Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado; PADIS, pain, agitation/sedation, delirium, immobility and sleep disruption; QTc, corrected QT interval; SD, standard deviation; SIAH, Sistema Integral de Administración Hospitalaria; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology; UNAM, Universidad Nacional Autónoma de México.

Ethics Approval and Consent to Participate

The Research, Ethics and Biosecurity Committees of Centro Médico Nacional "20 de Noviembre," ISSSTE, Mexico City, approved the study on 2 October 2024 (protocol RPI 643.2024). The committees waived the requirement for informed consent given the retrospective, records-based, standard-care-only design. Research access to clinical records began only after that approval; extraction and de-identification are dated in Section 2.7.

Human rights:

The study was conducted in accordance with the Declaration of Helsinki and the Reglamento de la Ley General de Salud en Materia de Investigación para la Salud (Título II, Capítulo I, Artículo 17), under which it was classified as research without risk ("sin riesgo").

Consent for Publication

Not applicable. This manuscript contains no individual person's identifiable data, images or clinical details. All records were de-identified before analysis, and no clinical-record codes appear in this report.

Availability of Data and Materials

The patient-level data are not publicly available due to confidentiality and ethical considerations but may be obtained from the corresponding author upon reasonable request, subject to applicable ethical approval, institutional authorization, and a data-use agreement, where required. The complete analysis script used to generate the reported numbers, tables, and figures is provided as Supplementary Material S2. No physical materials were used in this study. All figures are original and were generated de novo from the study's verified dataset; none has been published previously or reproduced from any copyrighted source.

Conflicts of Interest

The authors declare no competing interests, financial or non-financial, in relation to the work described in this manuscript.

Funding

No external funding was received for this research.

Authors' Contributions

Conceptualization, A.S.-R., A.D.-B. and M.G.O.-M.; methodology, A.S.-R., A.D.-B. and D.N.M.-S.; software, D.N.M.-S. and F.J.C.-A.; validation, D.N.M.-S., O.M.-L. and E.F.E.-d.-Á.; formal analysis, A.D.-B. and D.N.M.-S.; investigation, A.S.-R., A.D.-B., O.M.-L., E.F.E.-d.-Á. and F.J.C.-A.; resources, M.G.O.-M. and O.M.-L.; data curation, A.S.-R. and F.J.C.-A. (extraction and coding) and D.N.M.-S. and A.D.-B. (verification and reconciliation of the analytic dataset, including the duplicate and code-collision cases described in Section 2.4); writing—original draft preparation, A.S.-R., A.D.-B. and D.N.M.-S.; writing—review and editing, M.G.O.-M., O.M.-L., E.F.E.-d.-Á. and F.J.C.-A.; visualization, A.S.-R. and D.N.M.-S.; supervision,

M.G.O.-M.; project administration, A.D.-B. and M.G.O.-M. All authors have read and agreed to the submitted version of the manuscript.

Acknowledgements

The authors thank the staff of the Consultation-Liaison Psychiatry service of Centro Médico Nacional "20 de Noviembre," ISSSTE, for their support during data collection.

AI-declaration

During the preparation and revision of this work, the authors used a large language model (Claude, Anthropic) to assist with language editing, structural organisation of the text, the implementation of the statistical analysis code, and the generation of the figures from that code. The tool was used under author direction and produced no data, no results and no interpretation of its own. The authors credited with software (D.N.M.-S. and F.J.C.-A.) reviewed the analysis code line by line and re-ran it against the primary dataset; the authors credited with visualization (A.S.-R. and D.N.M.-S.) checked every figure against the values it plots. All analytic decisions, model specifications, interpretations and conclusions are those of the authors, who verified every result reported here against the primary dataset and take full responsibility for the integrity and accuracy of the article. No artificial intelligence tool is listed as, nor may be listed as, an author of this work.

Standards of Reporting

This observational cohort study was designed, conducted and reported in accordance with the STROBE statement for cohort studies [12]. The completed checklist is submitted as Supplementary Material S1.

Supportive/Supplementary Material

Supplementary Material S1 — STROBE checklist for cohort studies, documenting the location of each of the 22 reporting items within the manuscript. Supplementary Material S2 — the complete analysis script (Python) that reproduces every number, table and figure in this report from the de-identified dataset.

References

1. Wu CR, Chang KM, Tranyor V, Chiu HY. Global incidence and prevalence of delirium and its risk factors in medically hospitalized older patients: A systematic review and meta-analysis. Int J Nurs Stud. 2024;162:104959. doi:10.1016/j.ijnurstu.2024.104959.
2. Marcantonio ER. Delirium in hospitalized older adults. N Engl J Med. 2017;377(15):1456–66. doi:10.1056/NEJMcp1605501.
3. Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. Lancet. 2014;383(9920):911–22. doi:10.1016/S0140-6736(13)60688-1.
4. Nikooie R, Neufeld KJ, Oh ES, Wilson LM, Zhang A, Robinson KA, et al. Antipsychotics for treating delirium in hospitalized adults: a systematic review. Ann Intern Med. 2019;171(7):485–495. doi:10.7326/M19-1860.
5. Cruccioli MM, Carneiro PHT, Pitanga JFJ, Ferreira CC, Leão RC, Loures ML, et al. Haloperidol versus atypical antipsychotics for delirium in hospitalized and ICU patients: a systematic review and meta-analysis. Clin Neuropharmacol. 2025;48(5):145–150. doi:10.1097/WNF.0000000000000650.

6. Agar MR, Lawlor PG, Quinn S, Draper B, Caplan GA, Rowett D, et al. Efficacy of oral risperidone, haloperidol, or placebo for symptoms of delirium among patients in palliative care: a randomized clinical trial. JAMA Intern Med. 2017;177(1):34–42. doi:10.1001/jamainternmed.2016.7491.

7. Burton JK, Craig LE, Yong SQ, Siddiqi N, Teale EA, Woodhouse R, et al. Non-pharmacological interventions for preventing delirium in hospitalised non-ICU patients. Cochrane Database Syst Rev. 2021;7:CD013307. doi:10.1002/14651858.CD013307.pub2.

8. Jaqua EE, Nguyen VTN, Chin E. Delirium in older persons: prevention, evaluation, and management. Am Fam Physician. 2023;108(3):278–287.

9. Kim DH, Lee SB, Park CM, Levin R, Metzger E, Bateman BT, et al. Comparative safety analysis of oral antipsychotics for in-hospital adverse clinical events in older adults after major surgery: a nationwide cohort study. Ann Intern Med. 2023;176(9):1153-1162. doi:10.7326/M22-3021.

10. Timmons S, Bracken-Scally M, Chakraborty S, Gallagher P, Hamilton V, Begley E, et al. Psychotropic medication prescribing to patients with dementia admitted to acute hospitals in Ireland. Drugs Aging. 2023;40(5):461-472. doi:10.1007/s40266-023-01020-w.

11. Oviedo Lugo GF, García Martínez PA, Duarte Osorio A, Rincón CJ, López A, Chavarría R, et al. Incidence of confusional syndrome (delirium) in a Latin American university hospital. Rev Colomb Psiquiatr (Engl Ed). 2024;53(1):41-46. doi:10.1016/j.rcpeng.2022.01.002.

12. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med. 2007;147(8):573–7. doi:10.7326/0003-4819-147-8-200710160-00010.

13. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. Ann Intern Med. 1990;113(12):941–8. doi:10.7326/0003-4819-113-12-941.

14. Jones RN, Cizginer S, Pavlech L, Albuquerque A, Daiello LA, Dharmarajan K, et al. Assessment of instruments for measurement of delirium severity: a systematic review. JAMA Intern Med. 2019;179(2):231–239. doi:10.1001/jamainternmed.2018.6975.

15. Page VJ, Ely EW, Gates S, Zhao XB, Alce T, Shintani A, et al. Effect of intravenous haloperidol on the duration of delirium and coma in critically ill patients (Hope-ICU): a randomised, double-blind, placebo-controlled trial. Lancet Respir Med. 2013;1(7):515–23. doi:10.1016/S2213-2600(13)70166-8.

16. Girard TD, Exline MC, Carson SS, Hough CL, Rock P, Gong MN, et al. Haloperidol and ziprasidone for treatment of delirium in critical illness. N Engl J Med. 2018;379(26):2506–2516. doi:10.1056/NEJMoa1808217.

17. Andersen-Ranberg NC, Poulsen LM, Perner A, Wetterslev J, Estrup S, Hästbacka J, et al. Haloperidol for the treatment of delirium in ICU patients. N Engl J Med. 2022;387(26):2425–2435. doi:10.1056/NEJMoa2211868.

18. van der Vorst MJDL, Neefjes ECW, Boddaert MSA, Verdegaal BATT, Beeker A, Teunissen SCC, et al. Olanzapine versus haloperidol for treatment of delirium in patients with advanced cancer: a phase III randomized clinical trial. Oncologist. 2020;25(3):e570–e577. doi:10.1634/theoncologist.2019-0470.

19. Carayannopoulos KL, Alshamsi F, Chaudhuri D, Spatafora L, Piticaru J, Campbell K, et al. Antipsychotics in the treatment of delirium in critically ill patients: a systematic review and meta-analysis of randomized controlled trials. Crit Care Med. 2024;52(7):1087–1096. doi:10.1097/CCM.0000000000006251.

20. Zhao Y, Wang Q, Sun B, Li Y, Meng C, Miao G. Haloperidol for the treatment of delirium in ICU patients: a systematic review and meta-analysis. Eur J Med Res. 2025;30(1):147. doi:10.1186/s40001-025-02409-6.

21. Liu SB, Liu S, Gao K, Wu GZ, Zu G, Liu JJ. Olanzapine for the treatment of ICU delirium: a systematic review and meta-analysis. Ther Adv Psychopharmacol. 2023;13:20451253231152113. doi:10.1177/20451253231152113.

22. Veronese N, Solimando L, Bolzetta F, Maggi S, Fiedorowicz JG, Gupta A, et al. Interventions to prevent and treat delirium: an umbrella review of randomized controlled trials. Ageing Res Rev. 2024;97:102313. doi:10.1016/j.arr.2024.102313.

23. Wu YC, Tseng PT, Tu YK, Hsu CY, Liang CS, Yeh TC, et al. Association of delirium response and safety of pharmacological interventions for the management and prevention of delirium: a network meta-analysis. JAMA Psychiatry. 2019;76(5):526–535. doi:10.1001/jamapsychiatry.2018.4365.

24. Burry L, Mehta S, Perreault MM, Luxenberg JS, Siddiqi N, Hutton B, et al. Antipsychotics for treatment of delirium in hospitalised non-ICU patients. Cochrane Database Syst Rev. 2018;6:CD005594. doi:10.1002/14651858.CD005594.pub3.

25. Steinman MA. Alternative treatments to selected medications in the 2023 AGS Beers Criteria. J Am Geriatr Soc. 2025;73(9):2657–2677. doi:10.1111/jgs.19500.

26. Tomlinson EJ, Schnitker LM, Casey PA. Exploring antipsychotic use for delirium management. Drugs Aging. 2024;41(6):455–486. doi:10.1007/s40266-024-01122-z.

27. Ferrante LE, Chaudhuri D, Laiya Carayannopoulos K, Jain S, Tate JA, Álvarez-Espinoza E, et al. Society of Critical Care Medicine guidelines on caring for older adults in the ICU. Crit Care Med. 2026;54(5):1060–1072. doi:10.1097/CCM.0000000000007085.

28. Devlin JW, Skrobik Y, Gélinas C, Needham DM, Slooter AJC, Pandharipande PP, et al. Clinical practice guidelines for the prevention and management of pain, agitation/sedation, delirium, immobility, and sleep disruption in adult patients in the ICU. Crit Care Med. 2018;46(9):e825–e873. doi:10.1097/CCM.0000000000003299.

29. Oh ES, Fong TG, Hshieh TT, Inouye SK. Delirium in older persons: advances in diagnosis and treatment. JAMA. 2017;318(12):1161–74. doi:10.1001/jama.2017.12067.